Thai Journal of Mathematics : 45–58
Special Issue: Annual Meeting in Mathematics 2017

http://thaijmath.in.cmu.ac.th
ISSN 1686-0209

The Jackknife-Like Method for Assessing Uncertainty of Point Estimates for Bayesian Estimation in a Finite Gaussian Mixture Model

Kuntalee Chaisee

Department of Statistics, Faculty of Science, Chiang Mai University
Chiang Mai 50200, Thailand
e-mail : kuntalee.chaisee@cmu.ac.th

Abstract : In this paper, we follow the idea of using an invariant loss function in a decision theoretic approach for point estimation in Bayesian mixture models presented in [1]. Although using this approach the so-called label switching is no longer a problem, it is difficult to assess the uncertainty. We propose a simple and accessible way for assessing uncertainty using the leaving-out idea from the jackknife method to compute the Bayes estimates called jackknife-Bayes estimates, then use them to visualize the uncertainty of Bayesian point estimates. This paper is primarily related to simulation-based point estimation using Markov Chain Monte Carlo (MCMC) samples; hence the MCMC methods, in particular Gibbs sampling and Metropolis Hastings method are used to approximate the posterior mixture models. We also present the use of importance sampling in reduced posterior mixture distribution corresponding to the leaving-out observation.

Keywords : Jackknife method; finite Gaussian mixture models; MCMC methods.

2010 Mathematics Subject Classification : 47H09; 47H10.

1 Introduction

Bayesian approaches to mixture modeling are increasingly popular since the advent of simulation techniques, especially Markov chain Monte Carlo (MCMC) methods [2]. In Bayesian analysis of mixture models with known number of components, one of inferential difficulties is the problem so-called label switching in the MCMC output see [3] for a review. Using common practice for estimating parameters, the ergodic average of MCMC samples could be meaningless. One of the methods to deal with this problem suggested in [1] is to use a decision theoretic approach that computes parameter estimates using algorithms that aim to minimize the posterior expected invariant loss function. In general, point estimation is usually combined with interval estimation such as credible intervals or confidence intervals to show how reliable point estimates are. However, it is difficult to construct credible intervals or confidence intervals because of label switching.

In this paper, we denote a point estimate obtained from the decision theoretic approach by the Bayes estimate. We aim to provide a simple and accessible way for assessing uncertainty of Bayes estimates without dealing with label switching which usually occurs in Bayesian mixture models. Conception of uncertainty presenting in this paper is not the same as the interval estimation. We adopt the **leaving-out** idea from the jackknife method to compute the Bayes estimates called **jackknife-Bayes estimates** then use them to visualize the uncertainty of simulation-based Bayes estimates. What is the jackknife? In statistics, the jackknife method is often referred to as the nonparametric estimation of statistical error of a statistic of interest. It was introduced in [4] with the intention of reducing the bias of the sample estimate. It was developed further in [5] as a general approach for testing hypotheses and calculating confidence intervals using the assumption that the jackknife replicates are considered identically and independently distributed. The jackknife method is also known as a resampling method in which the basic concept is to estimate the precision of sample statistics based on removing data and then recalculating from subsets of available data known as the jackknife method or drawing randomly with replacement from a set of data points known as bootstrap method.

The paper is structured as follows. We initially study the use of MCMC methods for point estimation using a decision theoretic approach then implement the MCMC methods in R programming using simulated data. Then apply the jackknife-like method to compute the jackknife-Bayes estimates based on reduced-posterior distribution corresponding to the deleted observation. In this step, importance sampling is proposed to ease the computational requirement in MCMC simulation. Finally, we show uncertainty of Bayesian point estimation using the simulated jackknife-Bayes estimates.

2 Preliminaries

2.1 Bayesian Modeling on Finite Gaussian Mixture Distributions

Assume that $\boldsymbol{x} = (x_1, \ldots, x_n)$ is independent and identically distributed (i.i.d) observation from the distribution

$$f_{\boldsymbol{\theta}}(x_i) = \sum_{j=1}^{k} \omega_j \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left\{-\frac{(x_i-\mu_j)^2}{2\sigma_j^2}\right\}, \quad i = 1, \ldots, n, \tag{2.1}$$

where $\boldsymbol{\theta} = (\omega_1, \ldots, \omega_k, \mu_1, \ldots, \mu_k, \sigma_1^2, \ldots, \sigma_k^2)$, k is fixed and known and ω_j s are the weights satisfying $0 < \omega_j < 1$ and $\sum_{j=1}^{k} \omega_j = 1$. For convenience, we denote the parameters in (2.1) as follows: $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_k), \boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$, and $\boldsymbol{\sigma^2} = (\sigma_1^2, \ldots, \sigma_k^2)$. It is common to use the following priors in the Gaussian mixture model,

$$\begin{aligned}
\boldsymbol{\omega} &\sim \text{Dirichlet}(\delta_1, \ldots, \delta_k) \\
\mu_j &\sim \text{N}(\mu_0, \sigma_0^2) \quad \text{for } j = 1, \ldots, k, \text{independently} \\
\sigma_j^2 &\sim \text{InvGamma}(\alpha, \beta) \quad \text{for } j = 1, \ldots, k, \text{independently.}
\end{aligned}$$

Then prior probability distributions are

$$p(\boldsymbol{\omega}|k) = \frac{\Gamma(\delta_0)}{\Gamma(\delta_1)\ldots\Gamma(\delta_k)} \prod_{j=1}^{k} \omega_j^{\delta_j - 1}, \text{where } \delta_0 = \delta_1 + \ldots \delta_k, \tag{2.2}$$

$$p(\boldsymbol{\mu}|k) = \prod_{j=1}^{k} \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{1}{2\sigma_0^2}(\mu_j - \mu_0)^2\right\}, \tag{2.3}$$

$$p(\boldsymbol{\sigma^2}|k) = \prod_{j=1}^{k} \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\sigma_j^2\right)^{-(\alpha+1)} \exp\left\{\frac{-\beta}{\sigma_j^2}\right\}. \tag{2.4}$$

In mixture modeling, latent variables are often used to represent sub-populations where population membership is not known but is inferred from the data [6]. We define the variable $\boldsymbol{z} = (z_1, \ldots, z_n)$ as the latent variable used for allocating the components in the mixture model. We label the component j if $z_i = j$, $j \in \{1, \ldots, k\}$ for the ith observation. Suppose each z_i is independently drawn from the distribution such that

$$\text{Prob}\{z_i = j\} = \omega_j \qquad \text{for } j = 1, \ldots, k,$$

so the prior distribution of $\boldsymbol{z}$ is

$$p(\boldsymbol{z}|k, \boldsymbol{\omega}) = \prod_{i=1}^{n} \left(\omega_1 I_{A_1}(z_i) + \omega_2 I_{A_2}(z_i) + \ldots + \omega_k I_{A_k}(z_i)\right), \tag{2.5}$$

where I_{A_j} is an indicator function defined by

$$I_{A_j}(z_i) = \begin{cases} 1 & \text{if } z_i = j, \\ 0 & \text{otherwise,} \end{cases}$$

where $A_j = \{j\}$ for $j = 1, \ldots, k$. Define

$$n_j = \sum_{i=1}^{n} I_{A_j}(z_i), \quad \text{for } j = 1, \ldots, k, \tag{2.6}$$

then the prior distribution of $\boldsymbol{z}$ in equation (2.5) can be written as

$$p(\boldsymbol{z}|k, \boldsymbol{\omega}) = \prod_{j=1}^{k} \omega_j^{n_j}. \tag{2.7}$$

By using the latent variable $\boldsymbol{z}$ for allocation, we can express the likelihood function for the Gaussian mixture distribution in the form

$$l(\boldsymbol{x}|k, \boldsymbol{z}, \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma^2}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_{z_i}^2}} \exp\left\{-\frac{1}{2\sigma_{z_i}^2}(x_i - \mu_{z_i})^2\right\}. \tag{2.8}$$

By Bayes' Theorem applied to the priors in equations (2.2), (2.3), (2.4) and (2.7), and the likelihood function in equation (2.8), the posterior distribution can be expressed in the proportional distribution as follows:

$$\begin{aligned} \pi(\boldsymbol{\omega}, \boldsymbol{z}, \boldsymbol{\mu}, \boldsymbol{\sigma^2}|k, \boldsymbol{x}) &\propto p(\boldsymbol{\omega}|k) \times p(\boldsymbol{z}|k, \boldsymbol{\omega}) \times p(\boldsymbol{\mu}|k) \times p(\boldsymbol{\sigma^2}|k) \times l(\boldsymbol{x}|k, \boldsymbol{z}, \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma^2}) \\ &= \prod_{j=1}^{k} \omega_j^{\delta_j - 1} \times \prod_{j=1}^{k} \omega_j^{n_j} \times \prod_{j=1}^{k} \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{1}{2\sigma_0^2}(\mu_j - \mu_0)^2\right\} \\ &\times \prod_{j=1}^{k} (\sigma_j^2)^{-(\alpha+1)} \exp\left\{\frac{-\beta}{\sigma_j^2}\right\} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_{z_i}^2}} \exp\left\{-\frac{1}{2\sigma_{z_i}^2}(x_i - \mu_{z_i})^2\right\}. \end{aligned} \tag{2.9}$$

By using suitable priors, the conditional posterior distributions of the parameters $\boldsymbol{\omega}, \boldsymbol{\mu}$ and $\boldsymbol{\sigma^2}$ are the conjugate distributions. Therefore, we can use the standard MCMC method, Gibbs sampling to generate the samples of the parameter $\boldsymbol{\omega}, \boldsymbol{\mu}$ and $\boldsymbol{\sigma^2}$. Meanwhile, we use the Metropolis-Hastings method to generate the samples of the latent variable, $\boldsymbol{z}$ to allocate the components in the mixture distribution. We derive the conditional posterior distributions from the posterior distribution

(2.9) for each parameter as follows:

$$\pi(\boldsymbol{\omega}|k,\boldsymbol{z},\boldsymbol{\mu},\boldsymbol{\sigma^2},\boldsymbol{x}) \propto \prod_{j=1}^{k} \omega_j^{\delta_j+n_j-1}, \text{where } n_j\text{'s are from equation (2.6)}\,, \quad (2.10)$$

$$\pi(\boldsymbol{z}|k,\boldsymbol{\omega},\boldsymbol{\mu},\boldsymbol{\sigma^2},\boldsymbol{x}) \propto \frac{1}{\sqrt{2\pi\sigma_{z_i}^2}} \exp\left\{-\frac{1}{2\sigma_{z_i}^2}(x_i-\mu_{z_i})^2\right\}, \quad (2.11)$$

$$\pi(\mu_j|k,\boldsymbol{\omega},\boldsymbol{z},\boldsymbol{\sigma^2},\boldsymbol{x}) \propto \exp\left\{-\frac{1}{2b_j}(\mu_j-a_j)^2\right\}, \text{where} \quad (2.12)$$

$$b_j = \left(\frac{n_j}{\sigma_j^2}+\frac{1}{\sigma_0^2}\right)^{-1} \text{ and } a_j = \left(\frac{n_j}{\sigma_j^2}+\frac{1}{\sigma_0^2}\right)^{-1}\left(\frac{\mu_0}{\sigma_0^2}+\frac{n_j\bar{x}_j}{\sigma_j^2}\right),$$

$$\pi(\sigma_j^2|k,\boldsymbol{\omega},\boldsymbol{z},\boldsymbol{\mu},\boldsymbol{x}) \propto \left(\sigma_{z_i}^2\right)^{-(\frac{n_j}{2}+\alpha+1)} \exp\left\{-\frac{1}{\sigma_{z_i}^2}(\beta+\sum_{\substack{i=1\\ z_i=j}}^{n}\frac{(x_i-\mu_{z_i})^2}{2})\right\}. \quad (2.13)$$

As a result, the conditional posteriors are

$$\boldsymbol{\omega}|k,\boldsymbol{z},\boldsymbol{\mu},\boldsymbol{\sigma^2},\boldsymbol{x} \sim \text{Dirichlet}(\delta_1+n_1,\ldots,\delta_k+n_k),$$
$$\mu_j|\boldsymbol{\omega},\boldsymbol{z},\boldsymbol{\sigma^2},k \sim \text{N}(a_j,b_j),$$
$$\sigma_j^2|k,\boldsymbol{\omega},\boldsymbol{z},\boldsymbol{\mu} \sim \text{InvGamma}(\frac{n_j}{2}+\alpha,\beta+\sum_{\substack{i=1\\ z_i=j}}^{n}\frac{(x_i-\mu_{z_i})^2}{2})$$

$$\text{Prob}\{z_i=j|k,\boldsymbol{\omega},\boldsymbol{\mu},\boldsymbol{\sigma^2},\boldsymbol{x}\} \propto \frac{1}{\sqrt{2\pi\sigma_{z_i}^2}} \exp\left\{-\frac{1}{2\sigma_{z_i}^2}(x_i-\mu_{z_i})^2\right\}.$$

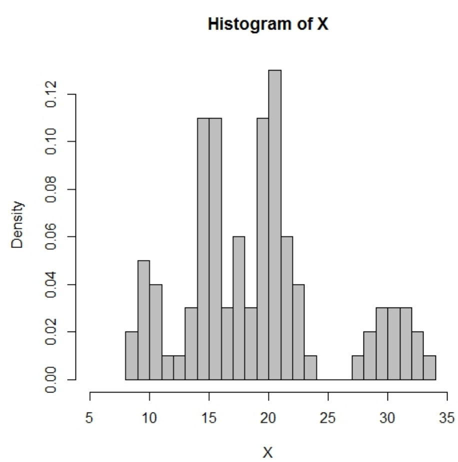


Figure 1: Histogram of simulated data from Gaussian normal mixture distribution with the number of components $k = 4$ with parameters defined in equation (2.14).

To present the use of the jackknife-Bayes estimates for assessing uncertainty, we use the simulated data in which the true parameters are assumed known. Consider the number of components $k = 4$. We independently simulate the data $\boldsymbol{x} = (x_1, \ldots, x_{100})$ from

$$X_i \sim 0.10N(10,1) + 0.25N(15,1) + 0.50N(20,2) + 0.15N(30,3), \tag{2.14}$$

for $i = 1, \ldots, 100$. Therefore, the simulated data $\boldsymbol{x} = (x_1, \ldots, x_{100})$ shown in Figure 1 is i.i.d from the distribution

$$f_{\boldsymbol{\theta}}(x_i) = \sum_{j=1}^{4} \omega_j \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left\{-\frac{(x_i-\mu_j)^2}{2\sigma_j^2}\right\}, \quad i = 1, \ldots, 100, \tag{2.15}$$

where

$$\begin{aligned} \boldsymbol{\omega} &= (0.10, 0.25, 0.50, 0.15), \\ \boldsymbol{\mu} &= (10, 15, 20, 30), \\ \boldsymbol{\sigma^2} &= (1, 1, 2, 3). \end{aligned}$$

Proceed the Bayesian mixture modeling explained earlier and use priors with the following hyper-parameters;

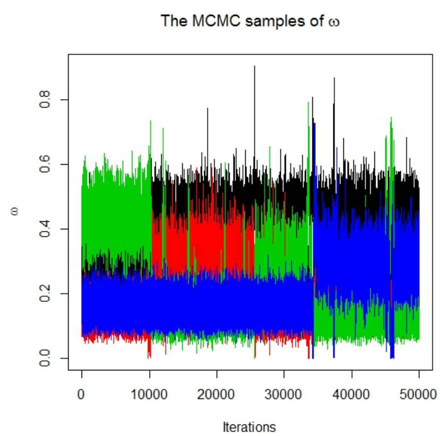
$$\begin{aligned} &\boldsymbol{\omega} \sim \text{Dirichlet}\,\underbrace{(1, \ldots, 1)}_{k} \\ &\mu_j \sim \text{N}(0, 100) \quad \text{for } j = 1, \ldots, k, \text{independently}, \\ &\sigma_j^2 \sim \text{InvGamma}(0.01, 0.01) \quad \text{for } j = 1, \ldots, k, \text{independently}. \end{aligned}$$

We simulate the MCMC samples of parameter $\boldsymbol{\theta} = (\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma^2})$ shown in Figure 2.

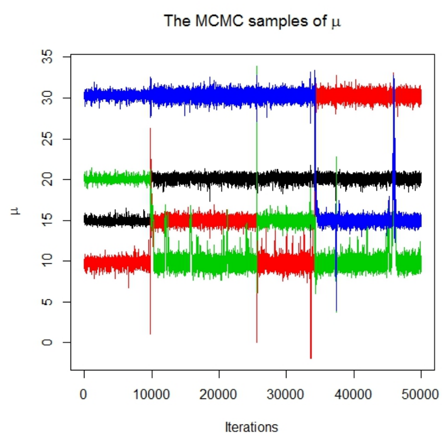
2.2 Bayesian Point Estimation Using Loss Function

In a Bayesian framework, parameter estimation can be done via decision theoretic approach. Typically, we need to specify a loss function, $L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$ which represents the loss incurred by estimating $\boldsymbol{\theta}$ with the estimate $\hat{\boldsymbol{\theta}}$ in the parameter space Θ. As a result, the point estimate will be to choose the value of $\hat{\theta}$ which minimizes the expected loss function with respect to the posterior distribution, $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ denoted by the Bayes estimate. One of well known loss functions is the quadratic loss function where the Bayes estimate is the posterior mean. In practice, the loss function is specified by the decision maker and can be complex, hence it is difficult derive the Bayes estimate related to the loss function analytically. Using numerical methods or simulation is more plausible for those loss functions. Consider the integrated squared difference,

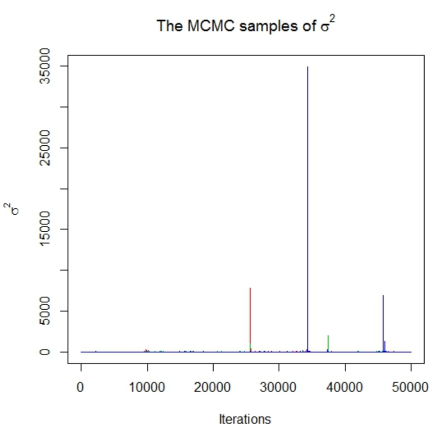
$$L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \int_{\mathcal{R}} (f_{\hat{\boldsymbol{\theta}}}(y) - f_{\boldsymbol{\theta}}(y))^2 dy, \tag{2.16}$$



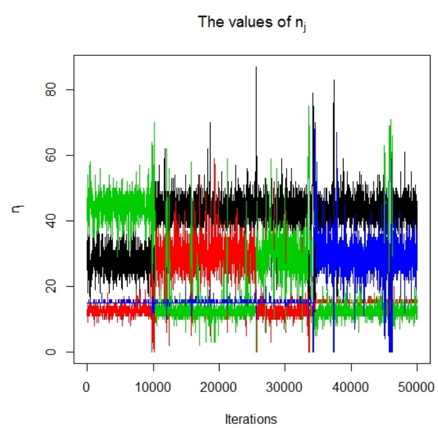
(a) $\boldsymbol{\omega}$



(b) $\boldsymbol{\mu}$



(c) $\boldsymbol{\sigma^2}$



(d) n_j for $j = 1, 2, 3$ and 4

Figure 2: The trace plots of MCMC samples of parameter $\boldsymbol{\theta} = (\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma^2})$ and n_j for $j = 1, 2, 3$ and 4

where $f_{\hat{\boldsymbol{\theta}}}(y)$ denotes the density of the Gaussian mixture model in equation (2.1). As a result, the integrated squared loss function (2.16) is label invariant. The Bayes estimate of the parameter denoted by $\boldsymbol{\theta}^*$ is the minimizer of the posterior expected loss. That is the value of $\boldsymbol{\theta}^*$ could be found by solving the minimization problem,

$$\hat{\boldsymbol{\theta}}^* = \arg\min_{\hat{\boldsymbol{\theta}}} \mathbb{E}_\pi \left[L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) \right]. \tag{2.17}$$

Given a potential $\hat{\boldsymbol{\theta}}$, we may define the posterior expected loss of a by averaging the loss function over the unknown parameter:

$$\begin{aligned}
\mathbb{E}_\pi \left[L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) \right] &= \int_{\boldsymbol{\theta}} L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) \pi(\boldsymbol{\theta} | k, \boldsymbol{z}, \boldsymbol{x}) d\boldsymbol{\theta} \\
&= \int_{\boldsymbol{\theta}} \int_{\mathcal{R}} ((f_{\hat{\boldsymbol{\theta}}}(y) - f_{\boldsymbol{\theta}}(y))^2 dy) \pi(\boldsymbol{\theta} | k, \boldsymbol{z}, \boldsymbol{x}) d\boldsymbol{\theta}
\end{aligned}$$

By using Bayesian sampling to obtain a Markov chain $\boldsymbol{\theta}^i = (\boldsymbol{\omega}^i, \boldsymbol{\mu}^i, \boldsymbol{\sigma^2}^i)$ for $i = 1, \ldots, N$ which converges to the stationary distribution $\pi(\boldsymbol{\theta} | k, \boldsymbol{z}, \boldsymbol{x})$. As a result, we have the MCMC samples which lead to an approximation of the posterior quantity as

$$\mathbb{E}_\pi \left[L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) \right] \approx \sum_{i=1}^{N} \int_{\mathcal{R}} (f_{\hat{\boldsymbol{\theta}}}(y) - f_{\boldsymbol{\theta}}(y))^2 dy.$$

Specifically, assuming the order of integration being interchanged, the expected posterior loss function can be written as

$$\mathbb{E}_\pi \left[L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) \right] = \int_{\mathcal{R}} f_{\hat{\boldsymbol{\theta}}}(y)^2 \mathrm{d}y - 2 \int_{\mathcal{R}} f_{\hat{\boldsymbol{\theta}}}(y) \mathbb{E}_\pi [f_{\boldsymbol{\theta}}(y)] \mathrm{d}y + \int_{\mathcal{R}} \mathbb{E}_\pi [f_{\boldsymbol{\theta}}(y)^2] \mathrm{d}y \tag{2.18}$$

By using ergodic averaging, we approximate

$$\mathbb{E}_\pi [f_{\boldsymbol{\theta}}(y)] \approx \frac{1}{N} \sum_{i=1}^{N} f_{\boldsymbol{\theta}^i}(y), \tag{2.19}$$

$$\mathbb{E}_\pi [f_{\boldsymbol{\theta}}(y)^2] \approx \frac{1}{N} \sum_{i=1}^{N} f_{\boldsymbol{\theta}^i}(y)^2. \tag{2.20}$$

We use simulated annealing as the numerical minimization technique to obtain $\hat{\boldsymbol{\theta}}^*$. For more details of simulated annealing algorithm and annealing schedule, see [7]. The simulation based Bayes estimates of the parameter $\boldsymbol{\theta} = (\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma^2})$ are shown in Table 1.

Table 1: Bayes estimates of parameter $\boldsymbol{\theta} = (\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\sigma^2})$

Parameter	True parameter	Bayes estimate
$(\omega_1, \omega_2, \omega_3, \omega_4)$	(0.10, 0.25, 0.50, 0.15)	(0.128, 0.288, 0.430, 0.154)
$(\mu_1, \mu_2, \mu_3, \mu_4)$	(10, 15, 20,30)	(9.79, 14.94, 20.11, 30.29)
$(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2)$	(1,1,2,3)	(1.465, 1.162, 2.345, 2.881)

3 Main Results

3.1 Jackknife-Like Method and Importance Sampling

The jackknife method was proposed by [4] and [5] and developed further as an approach for testing hypotheses and calculating confidence interval (see also [8]). The jackknife method is also known as a resampling method for variance and bias estimation. The basic concept of jackknife resampling is to estimate the precision of sample statistics based on removing data and then recalculate from subsets of available data. In this work, we only adopt the idea of leaving out observations (jackknifing) and recalculate the Bayes estimates. That is we compute Bayes estimates with respect to the reduced posterior distributions related to the deleted observation points. To avoid confusion, we call the Bayes estimate obtained from the reduced posterior distributions, the jackknife-Bayes estimates. Computing the jackknife-Bayes estimates can be computationally expensive as we have to use n different sets of the MCMC samples corresponding to the $\boldsymbol{x_i}$-delete posterior distributions, denoted by $\boldsymbol{x}_{(-i)}$ for $i = 1, \ldots, n$. We propose the use of importance sampling to make computation less expensive. To do so, we reuse MCMC samples generated from the full posterior distribution π which is regarded as the instrumental distribution in importance sampling. Therefore, we express the expected loss function with respect to each $\boldsymbol{x_i}$-delete posterior distribution as the expected loss function with respect to the (full) posterior distribution, $\pi(\boldsymbol{\theta}|\boldsymbol{x}) = \dfrac{p(\boldsymbol{\theta})l(\boldsymbol{x}|\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} p(\boldsymbol{\theta})l(\boldsymbol{x}|\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}}$ as follows:

$$\begin{aligned}\mathbb{E}_{\pi_{(-i)}}[L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})] &= \int L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{x}_{(-i)})\mathrm{d}\boldsymbol{\theta} \\ &= \int L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})\frac{\pi(\boldsymbol{\theta}|\boldsymbol{x}_{(-i)})}{\pi(\boldsymbol{\theta}|\boldsymbol{x})}\pi(\boldsymbol{\theta}|\boldsymbol{x})\mathrm{d}\boldsymbol{\theta},\end{aligned}$$

$$\begin{aligned}\mathbb{E}_{\pi_{(-i)}}[L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})] &= \int L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})\frac{p(\boldsymbol{\theta})l(\boldsymbol{x}_{(-i)}|\boldsymbol{\theta})/m_{-i}}{p(\boldsymbol{\theta})l(\boldsymbol{x}|\boldsymbol{\theta})/m}\pi(\boldsymbol{\theta}|\boldsymbol{x})\mathrm{d}\boldsymbol{\theta} \\ &= \frac{m}{m_{-i}}\int L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})w_i\pi(\boldsymbol{\theta}|\boldsymbol{x})\mathrm{d}\boldsymbol{\theta} = \frac{m}{m_{-i}}\mathbb{E}_{\pi}[L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})w_i], \qquad (3.1)\end{aligned}$$

where

$$w_i = \frac{l(\boldsymbol{x}_{(-i)}|\boldsymbol{\theta})}{l(\boldsymbol{x}|\boldsymbol{\theta})} = \frac{1}{l(x_i|\boldsymbol{\theta})} \qquad \text{if } x_i, i = 1, \ldots, n \text{ are independent,} \tag{3.2}$$

and

$$\begin{aligned} m_{-i} &= \int p(\boldsymbol{\theta}) l(\boldsymbol{x}_{(-i)}|\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \\ m &= \int p(\boldsymbol{\theta}) l(\boldsymbol{x}|\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}. \end{aligned}$$

Consider

$$\begin{aligned} m_{-i} &= \int \frac{p(\boldsymbol{\theta}) l(\boldsymbol{x}_{(-i)}|\boldsymbol{\theta})}{\pi(\boldsymbol{\theta}|\boldsymbol{x})} \pi(\boldsymbol{\theta}|\boldsymbol{x}) \mathrm{d}\boldsymbol{\theta} \\ &= \int \frac{p(\boldsymbol{\theta}) l(\boldsymbol{x}_{(-i)}|\boldsymbol{\theta})}{p(\boldsymbol{\theta}) l(\boldsymbol{x}|\boldsymbol{\theta})/m} \pi(\boldsymbol{\theta}|\boldsymbol{x}) \mathrm{d}\boldsymbol{\theta} \\ &= \int \frac{l(\boldsymbol{x}_{(-i)}|\boldsymbol{\theta})}{l(\boldsymbol{x}|\boldsymbol{\theta})/m} \pi(\boldsymbol{\theta}|\boldsymbol{x}) \mathrm{d}\boldsymbol{\theta}. \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{m}{m_i} &= \frac{1}{\int \frac{l(\boldsymbol{x}_{(i)}|\boldsymbol{\theta})}{l(\boldsymbol{x}|\boldsymbol{\theta})} \pi(\boldsymbol{\theta}|\boldsymbol{x}) \mathrm{d}\boldsymbol{\theta}} \\ &= \frac{1}{\int w_i \pi(\boldsymbol{\theta}|\boldsymbol{x}) \mathrm{d}\boldsymbol{\theta}} \\ &= \frac{1}{\mathbb{E}_\pi[w_i]}. \end{aligned} \tag{3.3}$$

From equations (3.1) and (3.3), we therefore estimate $\mathbb{E}_{\pi_{(i)}}[L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})]$ by using $\frac{\mathbb{E}_\pi[L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) w_i]}{\mathbb{E}_\pi[w_i]}$. We could use the same MCMC samples $\boldsymbol{\theta}^j$ for $j = 1, \ldots, N$ which are used in estimation $\mathbb{E}_\pi[L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})]$ to obtain $\mathbb{E}_\pi[L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) w_i]$ and $\mathbb{E}_\pi[w_i]$ by the following approximations

$$\mathbb{E}_\pi[L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) w_i] \approx \frac{1}{N} \sum_{j=1}^{N} L(\boldsymbol{\theta}, \boldsymbol{\theta}^j) w_i^j,$$

and

$$\mathbb{E}_\pi[w_i] \approx \frac{1}{N} \sum_{j=1}^{N} w_i^j,$$

where

$$w_i^j = \frac{l(\boldsymbol{x}_{(i)}|\boldsymbol{\theta}^j)}{l(\boldsymbol{x}|\boldsymbol{\theta}^j)}, \quad \text{for } i = 1, \ldots, n, j = 1, \ldots, N.$$

As a result, the expected loss function with respect to the x_i-delete posterior distribution in equation (3.1) can be computed by reusing the generated MCMC samples from the full posterior distribution π using the weight in equation (3.2). Importance sampling makes the estimation of the expected loss function relatively cheaply because we do not need to generate MCMC samples corresponding to the n x_i-delete posterior distributions. Nonetheless, computing the jackknife-Bayes estimates could still be expensive because of choices of minimization methods.

3.2 Simulation Results of Jackknife-Bayes Estimates

In this section, we present the simulation results of jackknife-Bayes estimates corresponding to the x_i-deleted points for $i = 1, \ldots, 100$. The jackknife-Bayes estimates of parameters $\boldsymbol{\omega}$, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma^2}$ are shown in Figures 3 - 5. Similarly to the trace plots, black, red, green and blue points represent the first, second, third and forth components of the jackknife-Bayes estimates of each parameter, respectively. Meanwhile, we also use black, red, green and blue horizontal lines to indicate the true values of the parameters corresponding to the true components. Due to the label switching problem, the order of the components of the estimated parameters from the simulation is not the same as the components of the true values.

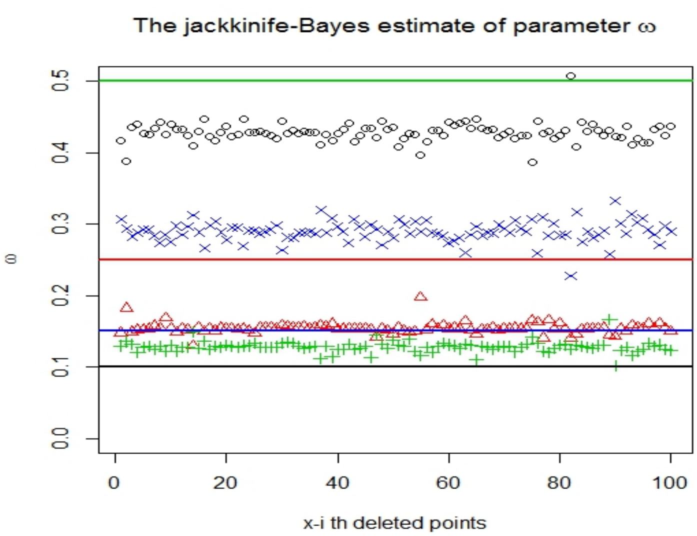


Figure 3: The jackknife-Bayes estimates of $\boldsymbol{\omega}$

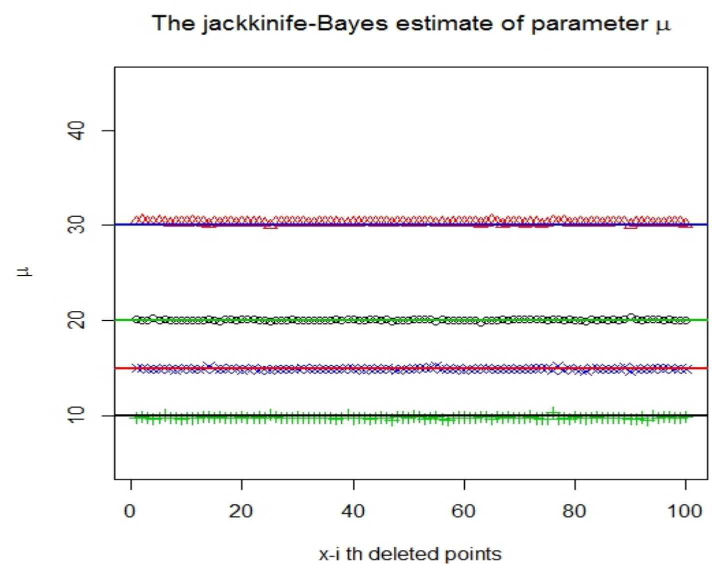


Figure 4: The jackknife-Bayes estimates of $\boldsymbol{\mu}$

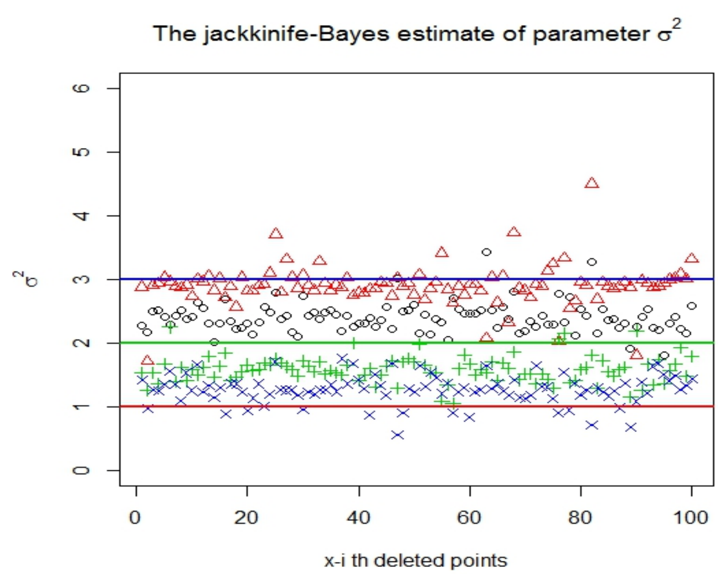


Figure 5: The jackknife-Bayes estimates of $\boldsymbol{\sigma^2}$

Figure 3 shows the simulation of the jackknife-Bayes estimates of the parameter $\boldsymbol{\omega}$, where the true value is $(0.10, 0.25, 0.50, 0.15)$. It shows that the jackknife-Bayes estimates perform fairly good, especially for the component of value 0.10 and the component of value 0.15 as they are clustered compared with the component of value 0.25 and 0.50. As a result, we might say that they are more certain than the other two components. Furthermore, the components of value 0.25 is slightly overestimated while the component of value 0.50 is slightly underestimated. For the parameter $\boldsymbol{\mu}$ in Figure 4, it is better than the parameter $\boldsymbol{\omega}$ and it is the most certain compared to the other parameters because all jackknife-Bayes estimates are very close to each other and also close to the true values $(10, 15, 20, 30)$. For the parameter $\boldsymbol{\sigma^2}$ shown in 5, we might say that they the least certain as all components of the jackknife-Bayes estimates are quite scattered and not as close to the true parameter as in the parameter $\boldsymbol{\mu}$.

4 Conclusions and Discussions

Generally, uncertainty of Bayesian point estimation in a finite Gaussian mixture model where the number of component is fixed and known cannot be shown by common practice such as credible intervals because of the label switching in MCMC output. In this paper, we can show the uncertainty by using the jackknife-Bayes estimates. It is an alternative approach to show uncertainty of Bayes estimates. The main results show that using jackknife-Bayes estimates can visualize the uncertainty of Bayes estimates of the parameter without difficulty dealing with label switching in MCMC output. Moreover, we show that importance sampling gives jackknife-Bayes estimates without too much computational demanding. But remember what we show in this paper is rely on two main methods, the MCMC generating method and the minimization method. According to MCMC simulation, we found that different hyper-parameters in priors leads to very different MCMC output hence the estimates. Therefore, choice of hyper-parameters should be chosen cautiously. Minimization method also plays important role in the search of Bayes estimates. We chose to use simulated annealing as it is applicable in our algorithm although it might be computationally expensive.

Acknowledgements : I would like to thank the referees for the valuable comments and suggestions on the manuscript. This work was supported by Faculty of Science, Chiang Mai University.

References

[1] G. Celeux, M. Hurn, R. Christian, Computational and inferential difficulties with mixture posterior distributions, Journal of the American Statistical Association 95 (2000) 957-970.

[2] C.P. Robert, G. Casella, A short history of Markov chain Monte Carlo: subjective recollections from incomplete data, Statistical Science 26 (2011) 102-115.

[3] A. Jasra, C.C. Holmes, D.A. Stephen, Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling, Statistical Science (2005) 50-67.

[4] M. Quenouille, Approximate tests of correlation in time series, Journal of the Royal Statistical Society. Series B (Methodological) 11 (1949) 68-84.

[5] J.W. Tukey, Bias and confidence in not-quite large samples, Annals of Mathematical Statistics 29 (1958) 614.

[6] G.J. McLachlan, D. Peel, Finite Mixture Models, John Wiley & Sons, New York, 2000.

[7] B.J.P. Claude, Convergence theorems for a class of simulated annealing algorithms on R^d, Journal of Applied Probability 29 (1992) 885-895.

[8] R.G. Miller, The Jackknife - a Review, Biometrika 61 (1974) 1-15

(Received 18 April 2017)
(Accepted 30 July 2017)